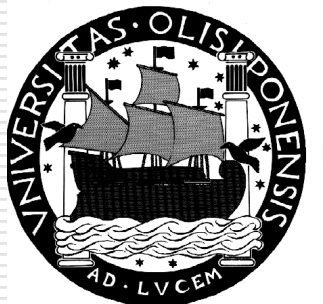


WP3 - Epiwork's Epidemic Marketplace

Mario Silva, Fabrício A. B. Silva, Francisco Couto, Luís Filipe Lopes, Patrícia Sousa, João Zamite, Hugo Ferreira
University of Lisbon, LASIGE - FFCUL



18 November - Epiwork Meeting
Torino

Outline

- ❑ Epiwork's WP 3 – Epidemic Marketplace
- ❑ Platform Requirements
 - Epidemic Marketplace General Architecture
- ❑ Epidemic Marketplace – Initial Deployment
- ❑ Epidemic Marketplace –Future Developments

Epidemic Marketplace

- A **data integration platform** where epidemiological data can be stored, managed and made available to investigators, fostering collaboration.
- The EM has four main components (all currently under development):
 1. Data Collector
 2. Repository
 3. Middleware Services
 4. Forum

EPIWork's WP3 - Epidemic Marketplace – Main Components

- 1. Repository with epidemic data sets and catalogue of epidemic datasources** containing the metadata describing existing datasets;
- 2. Forum** to publish information about their own data, seeking modellers to collaborate with, and/or to seek sources of data that could be of interest to their epidemiological modelling efforts;
- 3. Data collector**, which provides seamless integration of multiple heterogeneous data sources available on the Web;
- 4. Mediating software** that can automatically process queries for epidemiological data available from the information sources connected to the platform.

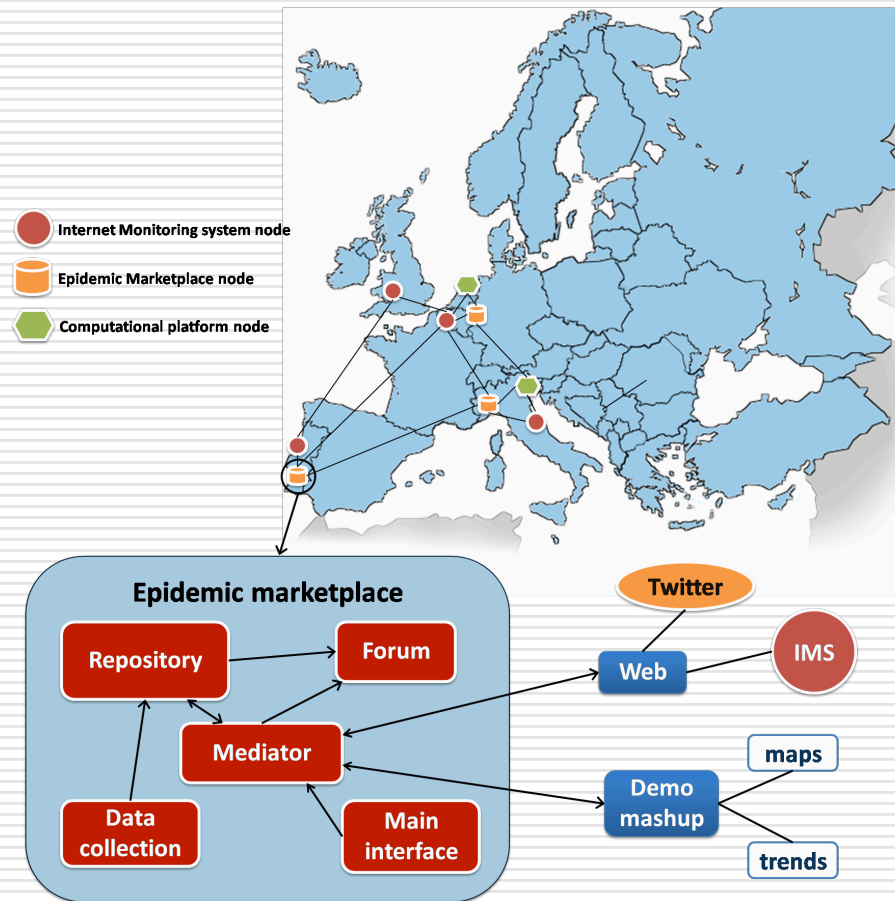
Platform Requirements

18 November - Epiwork Meeting
Torino

Epidemic Marketplace – Main Requirements

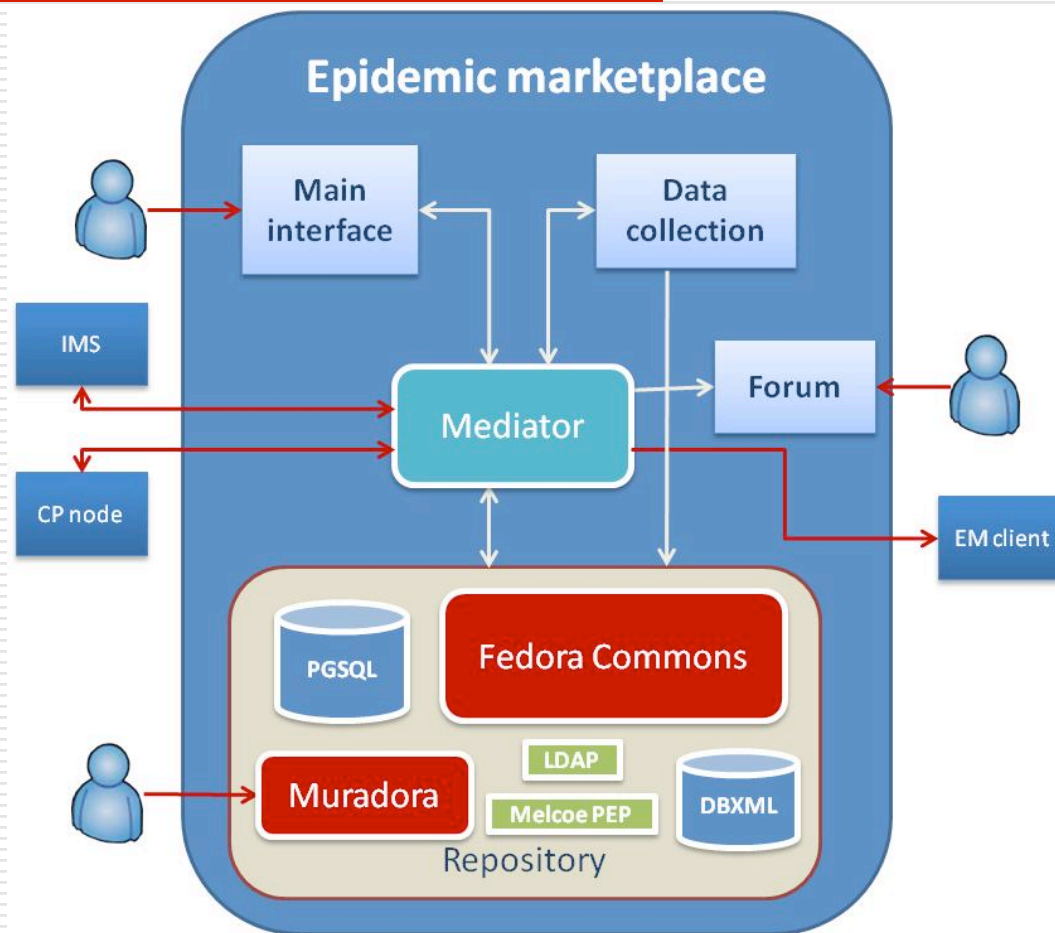
- ☐ **Support the sharing and management of epidemiological data sets**
 - Including anonymization?
- ☐ **Support the seamless integration of multiple heterogeneous data sources**
- ☐ **Support the creation of a virtual community for epidemic research**
- ☐ **Distributed architecture**
- ☐ **Support secure access to data**
- ☐ **Support data analysis and simulation in grid/cloud environments**
- ☐ **Workflow support**

Epidemic Marketplace Architecture



18 November - Epiwork Meeting
Torino

Epidemic Marketplace Architecture



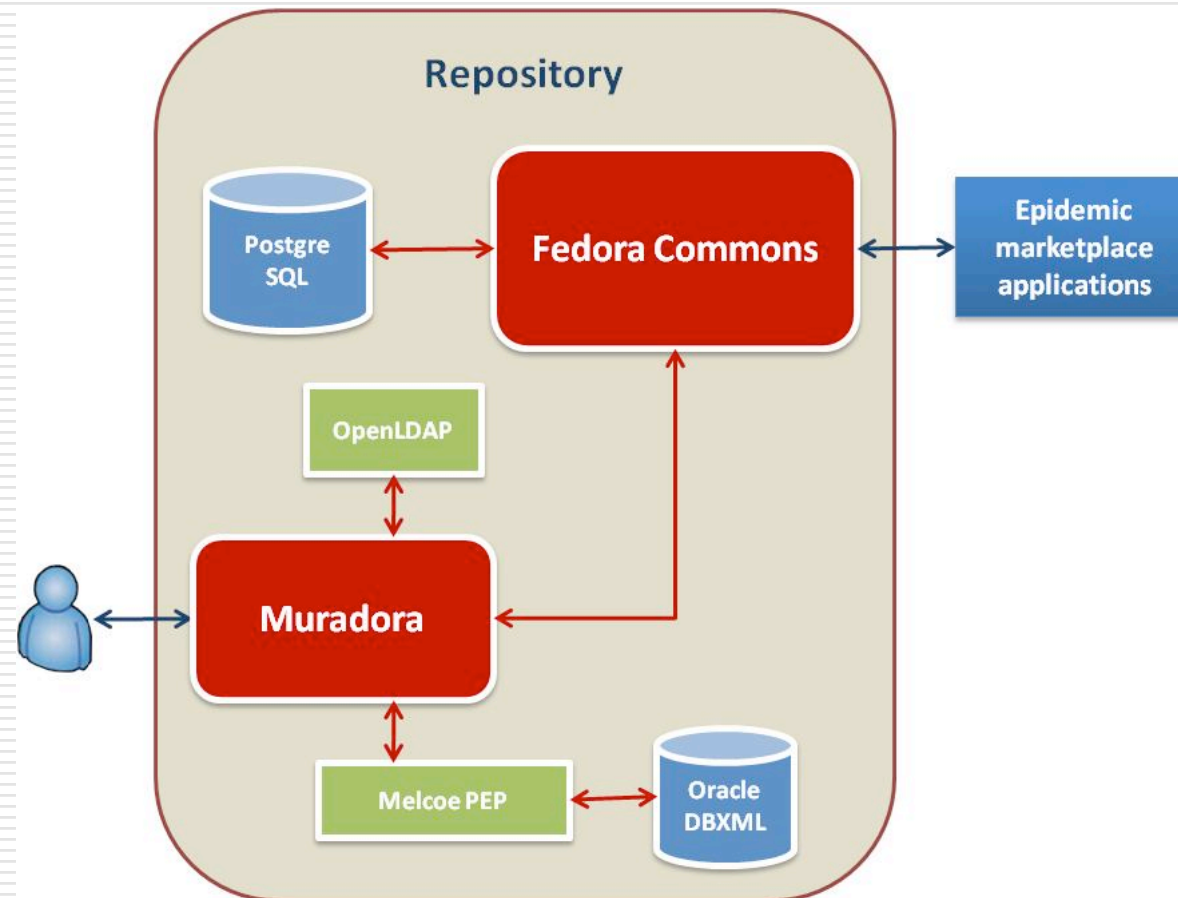
Epidemic Marketplace –Initial Deployment

18 November - Epiwork Meeting
Torino

Epidemic Marketplace – Initial Deployment

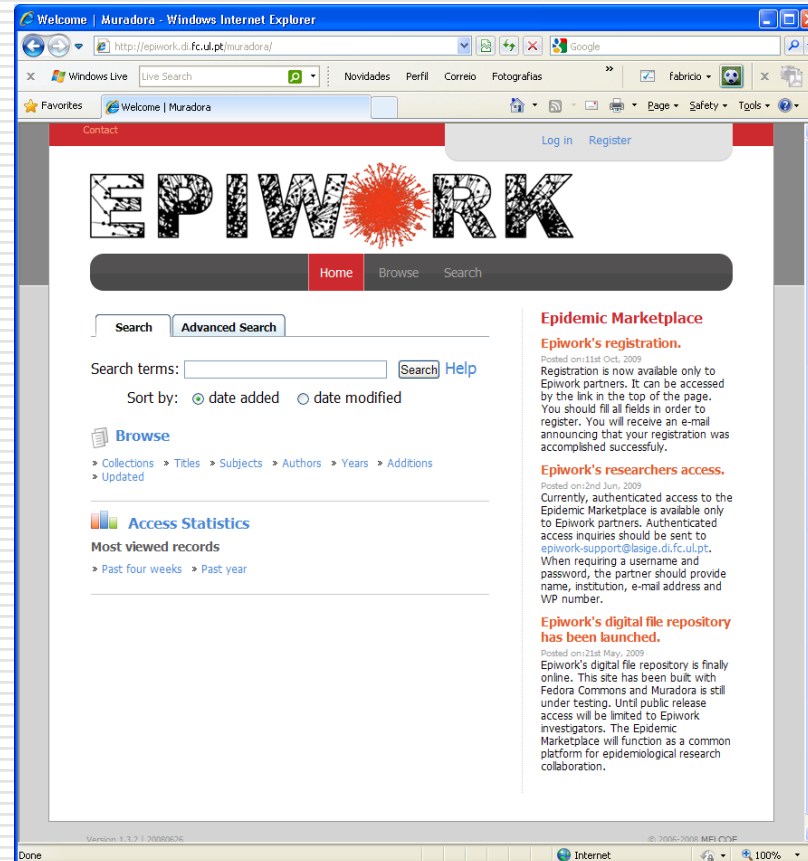
- ❑ Fedora used as Data Repository
- ❑ Front-end: Muradora
 - Dataset upload/download capability
- ❑ Dublin Core metadata support
- ❑ Browse and search capabilities
- ❑ Complex access control mechanisms

Epidemic Marketplace Repository– Initial Deployment



18 November - Epiwork Meeting
Torino

Epidemic Marketplace Repository – Initial Deployment



<http://epiwork.di.fc.ul.pt>

18 November – Epiwork Meeting
Torino

[Home](#)[Browse](#)[Search](#)[Submit](#)[Publish](#)[Portfolio](#)[Admin Tools](#)[Home](#)

Resource Metadata

Meta-data	
Title	Twitter dataset H1N1 + Portugal 4-6-2009
Creator	LASIGE node of the Epidemic Marketplace
Subject	twitter message dataset
Description	<p>This dataset contains Twitter messages containing the words H1N1 and Portugal collected between 16-5-2009 and 3-6-2009, Information is a 7 columns relation, containing the following data:</p> <p>Column 1- keyword 1 (disease)- H1N1 Column 2- Keyword 2 (location)- Portugal Column 3- Source (Twitter) Column 4- Author of the message (user id) Column 5- The message body (evidence) Column 6- score Column 7- date (day and hour)</p>
Publisher	Epiwork – http://www.epiwork.eu
Format	text/tab-separated-values
Language	English, Portuguese
Contributor	Luis F Lopes, Joao M Zamite, Bruno C Tavares, Francisco M Couto, Fabricio Silva, Mario J Silva
Relation	Luis F. Lopes, João M. Zamite, Bruno C. Tavares, Francisco M. Couto, Fabrício Silva and Mário J. Silva. (2009). Automated Social Network Epidemic Data Collector. INForum informatics symposium.
Source	http://epiwork.di.fc.ul.pt/collector/
Coverage	Spatial: Portugal, Temporal: 2009-5-16 to 2009-6-3
Rights	Creative Commons Attribution-ShareAlike (CC BY-SA), http://creativecommons.org/licenses/by-sa/3.0/

18 November – Epiwork Meeting
Torino

Epidemic Marketplace – Initial Deployment

- As the main site of the E.M., adequate hardware and network resources have been deployed at Lisbon
- Hardware platform
 - Two Dell PowerEdge SC1435 servers
 - Two Iomega Network Storage Servers
 - 12 TB total disk space

Epidemic Marketplace – Initial Deployment

☐ Network Infrastructure (1)

☒ 2Gb shared connection to FCCN

☐ FCCN is Responsible for the Portuguese infrastructure of advanced network for collaboration and communication in the fields of teaching and research

☐ Possibility of requesting additional 1Gb/s dedicated bandwidth increments, if project justifies it

Epidemic Marketplace – Initial Deployment

□ Network Infrastructure (2)

- 10Gb + 2.5 Gb connection between FCCN and the Géant Network (research & higher education European network)

Epidemic Marketplace – Data Collector

- ❑ **Data collector** -seamless integration of multiple heterogeneous data sources available on the Web
- ❑ First prototype retrieves data from twitter
- ❑ Main components
 - Data Storage
 - Web Crawler

Epidemic Marketplace – Data Collector

- Interacts directly with E.M.'s repository
 - Can publish epidemic data sets created from data retrieved through web crawlers
- First Prototype already available at <http://epiwork.di.fc.ul.pt/collector/index.htm>

Epidemic Marketplace – E.M. Forum

- **E.M. Forum** – The E.M. forum is already available.
 - Objective is to foster collaboration among partners and to disseminate information
 - Integrated with both the Repository and Data collector
 - **Based on which platform??**

Epidemic Marketplace – Future Developments

18 November - Epiwork Meeting
Torino

Epidemic Marketplace Mediator

- **Mediator**- This module will deploy a set of services available through a RESTful API that:
 - Exposes the data/metadata stored
 - Provides access to datasets
 - Interacts with other instances of the E.M.
 - Sends notifications to E.M.'s forum and external applications

Epidemic Marketplace Mediator

- ❑ External data sources can use API to download and upload datasets
- ❑ To be used by system developers/applications
- ❑ Compliant with OAI-PMH and OAI-ORE Standards

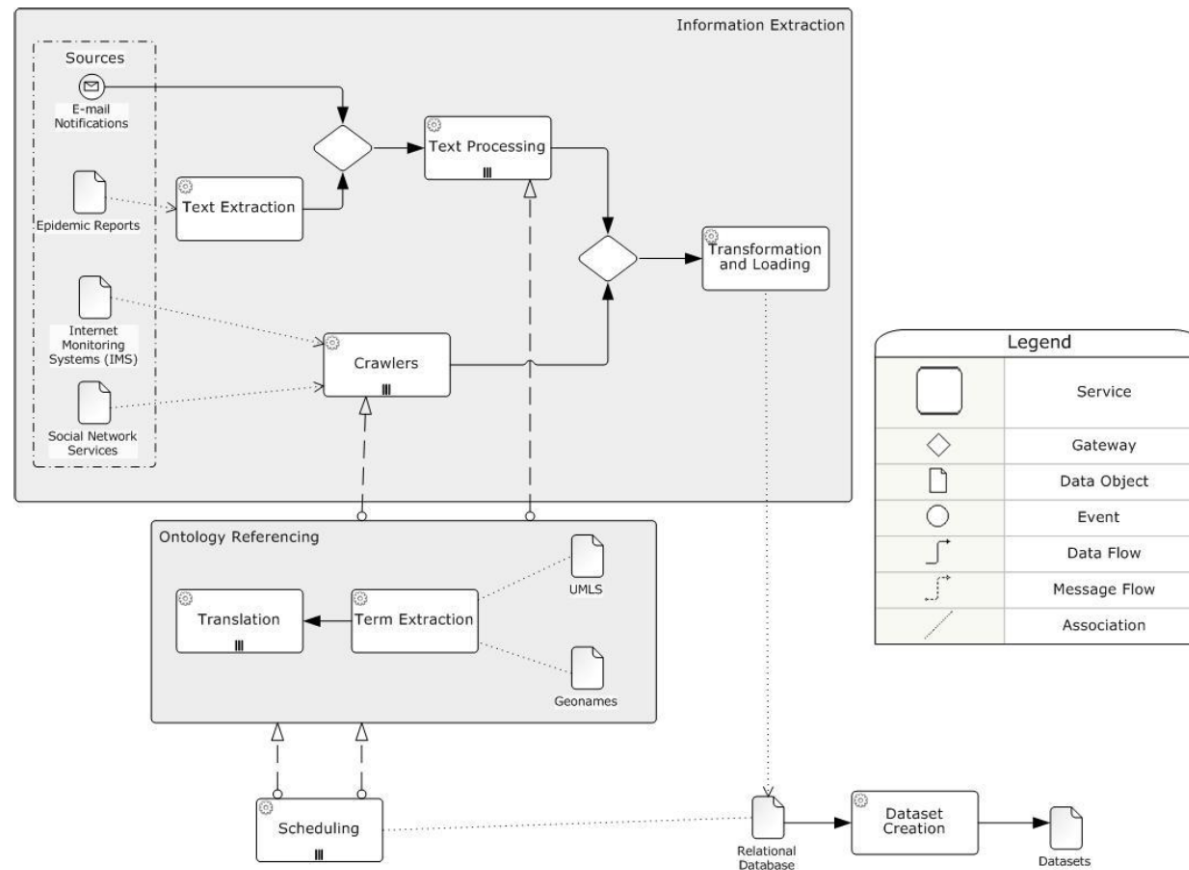
Mediator - APIs and Machine access

- ❑ PMH - Standard API for DC meta-data exchange
- ❑ ORE – SW style data constellations

Epidemic Marketplace – Multisource Data Collector

- Active and passive data collection
 - Multiple sources on the web
 - Data collections stored locally
- SOA, workflow, ontology support
- Easy to use service assembly interface, similar to Yahoo Pipes
 - BPEL Engine

Epidemic Marketplace – Multisource Data Collector



Evolution of E.M.'s Metadata Model

- ❑ Evolution of the original DC model
- ❑ Adapted for E.M. Datasets
 - We need your data!
- ❑ Ontology support
- ❑ Participation of Epiwork Partners in the creation and validation processes

Integration with WP2+WP4+WP5

- Data integration between WP3 and WP2+WP4+WP5 - several possibilities:
 - Upload/download of data sets
 - Streaming connection to WP3 Data Collector
 - Upload/download of metadata
- Built-in data/metadata search capability
 - Available through mediator services
 - Integrated access to data

Integration with WP2+WP4+WP5

- ❑ Workflow support
- ❑ Advanced authorization (single sign-on) and access control capabilities
- ❑ Enable integrated access to a comprehensive list of external data sources available through the Data Collector

Conclusion

- ❑ Initial Deployment of the Epidemic Marketplace available at **<http://epiwork.di.fc.ul.pt>**
- ❑ We need your data and your input!

Thank you!

epiwork@di.fc.ul.pt



18 November - Epiwork Meeting
Torino